

DanceGRPO: Unleashing GRPO on Visual Generation

Year: 2025 | Citations: 108 | Authors: Zeyue Xue, Jie Wu, Yu Gao, Fangyuan Kong, Lingting Zhu

Abstract

Recent advances in generative AI have revolutionized visual content creation, yet aligning model outputs with human preferences remains a critical challenge. While Reinforcement Learning (RL) has emerged as a promising approach for fine-tuning generative models, existing methods like DDPO and DPOK face fundamental limitations - particularly their inability to maintain stable optimization when scaling to large and diverse prompt sets, severely restricting their practical utility. This paper presents DanceGRPO, a framework that addresses these limitations through an innovative adaptation of Group Relative Policy Optimization (GRPO) for visual generation tasks. Our key insight is that GRPO's inherent stability mechanisms uniquely position it to overcome the optimization challenges that plague prior RL-based approaches on visual generation. DanceGRPO establishes several significant advances: First, it demonstrates consistent and stable policy optimization across multiple modern generative paradigms, including both diffusion models and rectified flows. Second, it maintains robust performance when scaling to complex, real-world scenarios encompassing three key tasks and four foundation models. Third, it shows remarkable versatility in optimizing for diverse human preferences as captured by five distinct reward models assessing image/video aesthetics, text-image alignment, video motion quality, and binary feedback. Our comprehensive experiments reveal that DanceGRPO outperforms baseline methods by up to 181\% across multiple established benchmarks, including HPS-v2.1, CLIP Score, VideoAlign, and GenEval. Our results establish DanceGRPO as a robust and versatile solution for scaling Reinforcement Learning from Human Feedback (RLHF) tasks in visual generation, offering new insights into harmonizing reinforcement learning and visual synthesis.